DOCUMENT RESUME

ED 062 606 AC 012 574

AUTHOR Woodley, Katheryn K.
TITLE Investigation of Test-Wiseness among C. L. U. Candidates.
PUB DATE Apr 72
NOTE 22p.

EDRS PRICE MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS *Cognitive Processes; *College Students; Control Groups; *Educational Research; Experimental Groups; Measurement Instruments; Objectives; Testing; *Test Results; *Test Wiseness

ABSTRACT

An attempt to help college students who felt that they understood the subject matter but couldn't pass the tests was made. For purposes of this study, test-wiseness (TW) is defined as a cognitive factor, one which is measurable and subject to change either through specific test experience or training in a test-taking strategy. The specific purpose was twofold: to gather empirical evidence about the level of test taking skills in the CLU population, and to develop an instructional program designed to improve these skills, if such a program were needed. In order to determine level of TW in the subjects studied, a test was constructed to measure selected-test-taking skills: (1) recognizing and eliminating similar options, (2) recognizing and eliminating absurd options, and (3) selecting an option which has a logical relationship with the stem. The students were divided into three groups: Program Experimental, Test Experimental, and Control. All were subject to pre-and post-testing. Because of the nature of the design of the present study, the norms for the CLU population on the TW Scale remain to be established. (Author/CK)

ED 062 606

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

INVESTIGATION OF TEST-WISENESS

AMONG C.L.U. CANDIDATES

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

Katheryn K. Woodley

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Katheryn K. Woodley
Zimmerman Adult Learning Laboratory
American College of Life Underwriters
April, 1972

ACO 12574

Definition of Test-Wiseness

Test-wiseness is a term which most researchers have probably heard or used, and often without a true understanding of the meaning of this fairly specific term. As a behavior, it is often confused with guessing or risk taking. As an explanation of test performance, it is often confused with bias or response sets, and very often is considered merely as part of undifferentiated error variance. To some people, the test-wise individual is seen as contributing to the unreliability of a test of knowledge, or interfering with the validity of a test of personality. Stanley has classified test-wiseness as one of the general and lasting characteristics of the individual in his analysis of sources of test variance (1971). He points out that it represents systematic variance, but that variation in the level of test-wiseness, when unrelated to the criterion of interest, will serve to reduce the validity of the test. He considers test-wiseness a real factor in almost any test score, since "freedom from emotional tension, shrewdness in guessing, and a keen eye for secondary and extraneous cues are likely to be useful in a wide range of tests" (1971, p.365).

Test wiseness is a construct, but has been given an operational definition so that it can be measured. That definition, as expressed by Oakland and Weilert (1971) is: "the ability to manifest test-taking skills which utilize the characteristics and formats of a test and/or test-taking situation in order to receive a score commensurate with the abilities being measured." Ebel and Damrin (1960) treated test-wiseness as a specific cognitive skill, capable of being developed through experience. They considered test-wiseness to be one of the four "bases" from which examinees could respond to objective test questions, clearly separating this ability from the other three--direct knowledge of content; response sets; and chance guessing.

The basic issue involved in test-wiseness seems to be one of determining

the extent to which a test validly descriminates on only those variables it was designed to measure (Oakland and Weilert, 1971). This recent statement is not at odds with the opinions expressed by most writers in this area (e.g. Thorndike, 1949; Ebel and Damrin, 1960; Vernon, 1962; Ebel, 1965; Millman and Setijadi, 1966). Several of these writers are of the opinion that, on a well-constructed test, a lack of test taking sophistication is a large source of error in measurement. Rather than viewing test-wiseness as insignificant or undesirable, the consensus seems to be that tests should be constructed with greater care and that people should be given training in how to take tests.

Based on a review of several studies, Millman, Bishop and Ebel (1965) outlined the test-wiseness principles, grouping them as either dependent on or independent of the test constructor or purpose. The following statement was included in their review. "There appears to be no systematic study of either the importance of test-wiseness or the degree to which it can be taught or measured" (1965, p. 707). The stated purpose of their analysis was to provide a framework within which future investigators could work, and they posed a series of questions for study. In spite of their excellent outline, very few studies since have focused directly on the problem. The terminology and framework have been increasingly adopted in the research that has been done, so that some benefit has been realized from the efforts of Millman and his colleagues. A skeleton diagram of their classification is shown in Appendix A.

One of the questions posed by Millman, Bishop and Ebel was whether or not test-wiseness can be taught. A number of recent studies have been directed at this question (e.g. Gibb, 1964; Moore, Schutz and Baker, 1966; Moore, 1968; Wahlstrom and Boersma, 1969; Slakter, Koehler and Hampton, 1970; Oakland and Weilert, 1971). In terms of the variety of learning experiences that have been designed, these studies reflect a rather broad based approach to providing

instruction in test-wiseness or related skills. There were varying degrees of success reported in these studies, and almost always, there was a criterion problem. Most of the programs and tests were designed for elementary or secondary school students, and no relevant studies focusing on an adult, non-college population were found. Although there have been several programs marketed which were designed to coach adults in dealing with specific tests (e.g. Civil Service, Armed Forces, CLEP), even the "popular" writers have not dealt with instruction in what could be termed general test-wiseness. On the basis of those studies dealing with other than adult populations, it would appear that the level of test-wiseness of an individual can be increased through training. Little evidence about persistence or the extent to which test-wiseness will generalize has been found.

Another question posed by Millman, Bishop and Ebel was related to the correlates of test-wiseness. No studies reflecting a comprehensive investigation of the correlates of test-wiseness were found, but several have focused on selected variables. For the most part, discussions of the personality correlates of test-wiseness have emphasized test anxiety, response sets, general mental ability, and risk-taking. The biographical variables receiving greatest attention have been sex and grade level (or age), largely because of the concentration of studies using elementary or secondary students.

The nature of the relationship between test-wiseness and anxiety has not been demonstrated. There is some evidence that familiarity with item types might lessen anxiety in a classroom situation, but whether or not this type of familiarity could be considered test-wiseness is debatable (Sassenrath, 1967). Although the idea that test sophistication and test anxiety are not compatible is generally accepted, empirical evidence is lacking. The importance of response sets for personality test scores has been will demonstrated in the literature (e.g. Cronbach,

1950; Bass, 1955; Couch and Keniston, 1960; Wevrick, 1962; Stricker, 1969). However, the concept is seen as relatively unimportant in multiple choice tests of achievement (Cronbach, 1950). In fact, the whole concept of test-wiseness appears to be different in personality and achievement tests.

Risk taking (on objective examinations) appears to be fairly consistent within a given test, but the relationship between this and test-wiseness remains to be demonstrated (Stone, 1962; Slakter, 1967). Slakter (1969) has suggested that a certain level of test-wiseness is essential before a subject can profit from taking risks. Although the feeling among researchers seems to be that general mental ability and test-wiseness are posibitvely correlated (e.g. Stanley, 1971), little real proof of this has been offered. In at least one study, the relationship between test-wiseness and general intelligence was not significant (Kreit, 1967). There is a similar paucity of research into the relationship of selected biographical characteristics to test-wiseness. Age has been shown to be positively correlated with test-wiseness for pre-school through high school students. No data on age or recency of test taking experience were available for adults.

It seems apparent that considerably more research into the nature of test-wiseness is needed. On the basis of a review of the recent literature, there would seem to be some agreement that people who are test-wise perform at a high level consistently, almost regardless of the type of test. There is evidence, however, that instructions in how to respond to specific types of items helps <u>specifically</u>. Stricker (1969) sees test-wiseness not as a broad, general ability, but rather as consisting of a set of "distinct and largely unrelated skills." Ebel and Damrin concluded that "insofar as 'test-taking' is a specific cognitive skill, it can, like any cognitive skill, be developed through experience. To the extent that differences in this skill are eliminated by adequate training, obtained

differences in test scores will provide better estimates of true difference between the capacities and abilities of individuals" (1960, p. 1511).

The Problem

The C.L.U. designation is awarded to qualified professionals in the insurance industry only upon successful completion of a series of ten achievement-type examinations. The examinations are prepared, administered and evaluated by the American College of Life Underwriters, a nonprofit organization which has been involved in this examination process for over 45 years. In addition to examination preparation, the College prepares a variety of study guides and learning aids to assist candidates in attaining the CLU designation. In addition to study and testing materials for the ten C.L.U. courses, several other adult education programs are offered. In all, the college currently serves approximately 60,000 students, administering examinations twice a year, in January and June.

The present study was initiated in response to a feeling among CLU candidates that they "understood the subject matter, but just couldn't pass the tests." This expression was in accord with a feeling among test developers and research staff at the College that the examination scores were probably contaminated somewhat by this population's lack of recent examination experience. This appeared a logical conclusion on the basis of the distributions of age and educational background of the CLU candidates. Approximately 35% of the candidates are 35 years of age or older when they begin their studies, and most have been away from an academic setting for quite a few years. It is entirely possible that a sizable number of new candidates have not taken an examination since high school or college. In some cases, it could have been 30 years since they've been faced with an achievement-type examination. Many insurance companies are beginning to require that their company officers have the CLU designation. Since the only way to obtain the designation is through successful completion of ten examinations, it would seem that

this population would have a strong incentive to improve their test taking abilities. Improvement in test taking skills should in turn improve the reliability and validity of the CLU examinations, decreasing the incidence of failure for reasons other than lack of knowledge.

For purposes of this study, we defined test-wiseness as a cognitive factor, one which is measurable and subject to change through either specific test experience or training in a test-taking strategy. Further, we made the assumptions that TW is complex, related to certain personality characteristics, and may be specific to the nature of the test, the test situation and the examiner. Based on these assumptions, our purpose was twofold: to gather empirical evidence about the level of test taking skills in the CLU population, and to develop an instructional program designed to improve these skills, if such a program were needed.

Test Development

In order to determine the level of TW in this population, it was necessary to construct a test to measure selected test-taking skills. Although some measures of TW had been developed as part of other studies, none were applicable to an adult population. The instrument developed for the measurement of TW consisted of 30 items, 10 items to measure each of three different TW skills. The test items had been designed so that each required the application of a specific test-taking strategy in order to arrive at the correct answer. Specifically, the test was designed to measure whether or not the examinee could arrive at the appropriate answer by: (1) recognizing and eliminating similar options; (2) recognizing and eliminating absurd options; and (3) selecting an option which has a logical relationship with the stem. Skills 1 and 2, referred to as "similar option" and "absurd option" skills, were included as deductive reasoning skills in the Millman, Bishop and Ebel classification, (1965), while skill 3, "stem option" was classified

as a cue using strategy. These specific skills were selected because of the cognitive processes implicit in their utilization, and because they seemed to bear a close relationship to the types of skills which might be needed on the CLU examinations. Further, it was possible to assess, directly, the ability to apply these strategies in a test situation.

The items, designed to measure application of the test-wiseness skills, were "nonsense" items. They were written as if measuring general knowledge but had no real right or wrong answers. It was necessary to use nonsense items instead of items reflecting any body of knowledge because of the variety of backgrounds within the CLU population. Correct responses could be made only through the application of a strategy or through chance guessing. The items were similar to those used by Slakter, in his test-wiseness measures (1970).

The TW items were all written by the author, then submitted to five judges for a content validity check. The judges were asked to sort the items into four stacks--one for each of the three TW skills with the fourth for items judged as not clearly reflecting any one of the skills. Items were retained only when there was unanimous agreement among the judges as to the nature of the TW skill measured. The items were then pretested on two adult populations.

The 30 TW items were imbedded in a test consisting of 30 legitimate, general knowledge test items. The legitimate items, reflecting several content areas, and utilizing item format similar to the TW items, were pretested on the same two adult populations. Only legitimate items of difficulty from 50% to 90% and with discrimination in the appropriate direction were retained for use in the final form of the test. The decision to imbed the TW items within a set of legitimate items was made to avoid the possibly debilitating effects of the examinees' either "giving up" or feeling overly threatened during the examination. Since the TW items were not content based, the examinees would have very little, if any,

positive reinforcement during the examination. It was hoped that some immediate positive reinforcement could be provided through the addition of legitimate items of fairly low difficulty. All test items were multiple choice, and written in the formats commonly used for vocabulary, arithmetic calculations and general knowledge type items. The items were organized within the test according to item type, rather than test-taking strategy, in the order given above. Except for the arithmetic calculations items which were all legitimate, TW and legitimate items were randomly ordered within the test sections.

Reliability was estimated for the total test and for each of the test-taking strategy subtests. Based on a sample of 104 CLU candidates, the Cronbach alphas shown in Table I were obtained.

TABLE I

TEST-WISENESS SCALE RELIABILITY

Subtest	Test Strategy	Number of Items	Alpha
Subtest I	Similar Option	10	0.44
Subtest II	Absurd Option	10	0.52
Subtest III	Stem Option	10	0.63
Total Test		30	0.73

Reliability will be estimated after each test revision, as well as with each new population tested.

Survey of TW in CLU Population

Early in the fall 1971 semester, the 30 item TW scale was administered to a total of 259 CLU students, enrolled in 15 classes. The classes were selected on the basis of geographic location and willingness of the teacher to participate in the project. Admittedly, this might not seem the most desirable means of class selection, but the need for controlled testing and possibly frequent in-person contact with the subjects rendered this the only feasible means of investigation. To allow for control on an outside criterion, only classes in Course One, "Individual Life and Health Insurance," were included. Class size ranged from 8 to 40, with a median of 15. Because the TW scale is somewhat transparent if the purpose is known, it was important that the students did not know why they were taking the test. This necessitated careful control during the test administration. All tests were administered in the regular classroom, during a class session, by someone from the College who had been given directions about the amount of information which could be transmitted to the subjects.

Biographical information about each student was collected during the fall test administration. Any students who were not in attendance on the day the test was given were not included in the sample, but a record of total class size was kept. The biographical information was to be used in the program evaluation phase of the study.

The overall range of scores on the test was 5 to 29. The descriptive statistics for each of the 15 classes are shown in Appendix B.

The results of the test administration indicated that there were some differences in the levels of TW, as measured by our test, in this population. The study had been set up so that if the need for an instructional program in TW seemed apparent the fall administration of the TW scale could serve as a pre-test for formative evaluation of this program. Since we were reasonably certain that

an instructional program would be beneficial to the CLU examinees, we decided to develop a program aimed directly at this population.

Test-Wiseness Program Development

The TW program developed for fall testing was devoted entirely to training people to respond to objective, multiple-choice type test items. It combined instruction and measurement in a workbook format, with diagnostic testing and prescribed branching built-in. The program was divided into four sections: an introduction, primarily aimed at anxiety reduction; an overview, in which the test-taking strategies were reviewed and examples provided; a diagnostic-branching section, requiring application of key strategies and providing specific instruction to program users as needed; and a final review test, sampling knowledge of principles and providing page references for review of questions answered incorrectly.

It is, therefore, a self-contained package of instruction, measurement and suggestions for review. Unlike most programs designed to teach test-taking, we did not focus on practice in the types of items used in the CLU examination. Although these items were used to illustrate some of the principles, the focus was on instruction in specific strategies. A total of eleven such strategies were included in the program. While the program was designed to provide instruction in most of the generally accepted test-taking strategies, it became apparent that the level of instruction needed was not the same for all skills. Specifically, while some of the skills clearly required proficiency at the application level, others seemed amenable to instruction at the knowledge level, with application skills assumed as a result of knowledge. All skills, even those taught at the application level, were first taught at the knowledge-recognition level. The following breakdown illustrates the treatment given to different skills:

STRATEGIES INCLUDED IN TEST-WISENESS PROGRAM	
Knowledge Level (only)	Application Level
1. Time using strategy 2. Error avoidance strategy 3. Guessing strategy 4. Conflicting options (Deductive) 5. Utilization of information given elsewhere to answer specific test items (Deductive) 6. Grammatical cues (Cue using) 7. Intent consideration strategy	1. Stem-option (Cue using) 2. Similar option (Deductive) 3. Absurd option (Deductive) 4. Specific Determiners (Cue using)

The final review, placed at the end of the program, covered all the skills and served as a final check on knowledge of test-taking strategies.

The completed program was made available to a sample of CLU candidates, for purposes of formative evaluation.

Program Evaluation

Since we were reasonably certain from the beginning that some type of instructional program would be developed, we decided to administer the TW Scale early enough in the semester that it could be used for program evaluation. As mentioned in an earlier section, the TW Scale was administered as a pre-test to a total of 259 CLU students enrolled in 15 Course One classes. Biographical information, collected during test administration, was used in matching class profiles to arrive at the experimental groupings, in keeping with the pre-test, post-test, control group design of the study. The information collected included: age, level of education achieved, years in the insurance field, number of years since taking an educational examination, number of CLU examinations previously taken, class size, and the experience level of the teacher. Class averages for

each of these variables, as well as for performance on the TW Scale were determined. These are shown in Appendix C.

Because of the nature of the study, matching was done on the basis of class profiles rather than on an individual student basis. The class averages were used to plot profiles and intuitive matching used to arrive at experimental group classification. As a result of matching on those variables typically or intuitively related to TW, equivalent contribution to error from the factors more likely affecting performance was assumed.

Based on the profiles, each of the classes was placed into one of three groups: Program Experimental, Test Experimental, or Control, with five classes in each group. As a result of the grouping, a total of 87 students were in the Program Experimental group; 92 in the Test Experimental group, and 80 in the Control group. The pre-test, post-test, Control group design of the study was thus enhanced with an additional "moderate intervention" group for purposes of information collection and added control. For the Program Experimental group, the test-wiseness program described in the previous section was used as the intervention. A battery of psychological tests was administered to the Test Experimental group shortly before the end of the semester. All three groups completed the same TW Scale as a post-test prior to completing the CLU examination for Course One. The following diagram illustrates the design.

RESEARCH DESIGN FOR PROGRAM EVALUATION

GROUP I Program Experimental	GROUP II Test Experimental	GROUP III Control
Test-Wiseness Scale (pre-test)	Test-Wiseness Scale (pre-test)	Test-Wiseness Scale (pre-test)
Test-Wiseness Program (intervention)	Test Battery A. Intelligence B. Anxiety C. Personality D. Biographical (intervention)	(no intervention)
Test-Wiseness Scale (post-test)	Test-Wiseness Scale (post-test)	Test-Wiseness Scale (post-test)
CLU Examination-I (outside criterion)	CLU Examination-I (outside criterion)	CLU Examination-I (outside criterion)

Criteria - Change in Test-Wiseness from pre- to post-test
- Reduction in variance, increase in mean on CLU examination
- Reliability of performance over time (consistency of performance on in-class tests)

The Test Experimental group served a dual purpose: it provided information about some of the correlates of TW without introducing possibly contaminating effects into the program evaluation phase of the study. It also provided us with some information about the effects of recent systematic and comprehensive testing on the level of TW. The test battery consisted of: the Advanced Mental Ability Test, the Gordon Personal Profile, the IPAT Anxiety Scale, the Personnel Data Questionnaire (a biographical information questionnaire) and the Multi-Aptitude Battery. Except for the Personnel Data Questionnaire which was mailed to the students for completion, these tests were administered in one setting. Participation in this phase had to be on an individual volunteer basis, as many of the classrooms were not available for our use other than during the regular class period. The

teaching schedules were such that the two hour battery could not be administered during a regular class meeting. Even though all students enrolled in Test Experimental classes were solicited, only about one fourth of them completed all the tests. It is possible that our inability to disclose the purpose of the research project until after the post-test was completed, together with a general negative attitude toward taking tests, were to blame for the low turnout in this generally "willing" audience. At present, all tests have been scored, but any discussion of the results must wait until further testing is completed. This part of the study will be repeated this spring, to provide greater insight into the nature of test-wiseness.

The performance of the three groups on both the TW post-test and the CLU examination will be compared in an effort to evaluate program effectiveness. Gain scores will be calculated for all three TW skills, to see if there is any difference in growth among the skills. The CLU examination, since its preparation is completely outside the control of this study, serves as an outside criterion. Since TW yields more consistent scores, or less error variance, we would expect the inter-individual reliability for the CLU examination to be highest for the Program Experimental group. This will be measured by comparing examination performance of students who have and students who have not received TW training. If the test-wiseness program is effective, we would also expect that the intra-individual response variability would be lower for the group given the test-wiseness treatment. This necessitates some measure of stability of performance over time and unfortunately the program was not available for distribution early enough in the semester to collect such data on the fall sample. To the extent that it is possible, records of in-class tests for future courses will be kept for those students using the program so that this can be ascertained. For the present study, a record of performance on in-class tests for all students was obtained.

Change in rank on these tests will be considered a measure of intra-individual response variability, and will be correlated with performance on the pre-test administration of the TW Scale, to see if there is a relationship in the expected direction. It is therefore used as further evidence of the validity of the test rather than as an indication of program effectiveness.

Theoretically, by comparing the Program Experimental with the Control group we can determine the effect of the program on both the measured level of test-wiseness and performance on an outside criterion. Comparing the Test Experimental group with the Control group will demonstrate the effect of systematic and comprehensive testing, again on both the measured level of test-wiseness and performance on an outside criterion. By examining the measures obtained from the Test Experimental group we can gain insight into the correlates of test-wiseness, with the possibility of future construct validity. Because of the attrition in the Test Experimental group, however, some of these comparisons will have to await replication.

Summary

The evaluation of the TW program as conducted for the fall sample was formative, designed to judge the difficulty and applicability of the materials for this population. It served to provide feedback about the effectiveness of this approach in improving the test-taking skills of this very specific population. The program is currently undergoing revision based upon the results of this evaluation. In addition to the incorporation of changes indicated by the fall testing, it is being expanded to include a section on essay and short answer completion type items. Eventually, research design will demand that summative evaluation of the program be carried out, but this is not anticipated until the January, 1973 examination period. Although a sizable sample will be available for study prior to the June, 1972 examinations, the focus will be on the

correlates of TW, further evaluation of the existing program, and initial testing of the new sections of the program.

In addition to continuing research with the CLU candidate population, similar testing-program studies will be completed on two additional adult populations. This expansion will give meaningful feedback about the TW Scale as well as whether or not the program is generalizable. The spring samples are both comprised of college students, but at distinctly different levels. One sample consists of three classes from a junior college, whose student body is made up of girls with histories of under-achievement or whose scholastic abilities are not sufficient for them to survive in a typical college situation. The other sample is made up of senior and graduate students in psychology from a major university. Plans for expanding to other adult "vocational" groups have been discussed, but will not be formulated until the results of the current investigations are analyzed.

Because of the nature of the design of the present study, the norms for the CLU population on the TW Scale remain to be established. Further testing should be directed toward determining the level of TW in the CLU, and in other populations, as projected from random samples. Further reliability estimates, continuous item refinement based on item analyses and normative data collection are planned as part of the College's ongoing research into this problem.

APPENDIX A

Test Wiseness Strategies *

I. Characteristics Dependent On Test Constructor Or Purpose

- A. Intent Consideration strategy
- B. Cue-using strategy
  1. Recognition of specific determiners
  2. Recognition of similarities between an option and an aspect of the stem
  3. Recognition of any consistent idiosyncrasies of the test constructor

II. Characteristics Independent of Test Constructor Or Purpose

- A. Time using strategy
- B. Error avoidance strategy
- C. Guessing strategy
- D. Deductive reasoning strategy
  1. Recognition of similar options
  2. Recognition of absurd options

* From: Millman Bishop & Ebel, 1965

APPENDIX B

TEST WISENESS PRE-TEST DESCRIPTIVE STATISTICS

Class	N	Mode	Median	Range	Mean	S.D.
1	32	16	19	10 to 29 (19)	18.5	4.3
2	21	20	21	15 to 28 (13)	21.05	2.84
3	14	22	20.5	16 to 27 (11)	20.5	3.31
4	11	18	21	15 to 27 (12)	20.27	3.53
5	17	20	1	5 to 29 (24)	17.47	5.78
6	13	17	18	11 to 25 (14)	18.31	3.74
7	10	16, 19	19	14 to 26 (12)	19.5	3.72
8	28	19	19	13 to 28 (15)	18.79	3.54
9	8	None	18.5	13 to 28 (15)	19.38	4.8
10	13	(16,19,22,25)	20	15 to 26 (11)	20.38	3.63
11	17	21	21	12 to 26 (14)	20.6	4.34
12	11	19, 20	17	7 to 22 (15)	15.54	4.97
13	15	20	21	18 to 28 (10)	21.87	2.68
14	40	19	19	13 to 27 (14)	18.85	3.72
15	9	18, 22	19	14 to 27 (13)	20.22	3.89

APPENDIX C

CLU CLASS STATISTICS

Class	Age (Median) Years	Education (N) High School	(N) Attend College	(N) College Grad.	(N) Adv'd. Study	Years Since Last Exam	CLU Exams Taken One (N)	More Than One (N)	Yrs. in Insurance (Mean)
1	31.5	9	9	10	4	2.8	0	1	3.7
2	32	9	9	3		2.1	1	7	6.4
3	27.5	4	1	3	6	2.1	1	2	4.9
4	35	7	1	2	1	3.1	0	1	4.1
5	36	10	4	3		2.6	0	3	6.3
6	39	7	3	3		3.1	0	0	8.4
7	34	8	1	1		2.9	2	3	7.8
8	33.5	13	5	4	6	3.6	1	2	5.3
9	33.5		5	2	1	2.5	1	1	4.4
10	37	2	3	4	4	2.8	1	3	8
11	31	3	3	10	1	2.4	4	0	6.4
12	26	4	2	4	1	2.4	0	1	5.8
13	27	1	4	9	1	2.3	0	0	4.2
14	30	8	14	14	4	3.7	3	o	5.7
15	34	2	2	5		1.8	0	4	7.4

REFERENCES

Bass, B.M. "Authoritarianism or Acquiesence?" Journal of Abnormal and Social Psychology, 1955, 51, 616-623.

Cronbach, L.J. "Further Evidence on Response Sets and Test Design," Educational and Psychological Measurement, 1950, 10, 3-31.

Couch, A. & Keniston, K. "Yeasayers and Naysayers: Agreeing Response Sets as a Personality Variable," Journal of Abnormal and Social Psychology, 1960, 60, 151-174.

Ebel, R.L. Measuring Educational Achievement. New Jersey: Prentice-Hall, 1965.

Ebel, R.L. & Damrin, D.E. "Tests and Examinations." In Chester W. Harris (ed.), Encyclopedia of Educational Research (3rd. ed.), New York: The Macmillan Company, 1960, 1502-1517.

Gibb, B.G. "Test-Wiseness as Secondary Cue Response," Unpublished doctoral dissertation, Stanford University, 1964.

Kreit, L. "The Effects of Test-Taking Practice on Pupil Test Performance," Unpublished doctoral dissertation, Indiana University, 1967.

Millman, J. & Setijadi. "A Comparison of the Performance of American and Indonesian Students on Three Types of Test Items," The Journal of Educational Research, 1966, 59, 273-275.

Millman, J., Bishop, C.H. & Ebel, R. "An Analysis of Test-Wiseness," Educational and Psychological Measurement, 1965, 25, 707-726.

Moore, J.C. "Manipulating the Effectiveness of a Self-Instructional Program," Journal of Educational Psychology, 1968, 59, 315-319.

Moore, J.C., Schutz, R.E. & Baker, R.L. "The Application of a Self-Instructional Technique to Develop a Test-Taking Strategy," American Educational Research Journal, 1966, 3, 13-17.

Oakland, T. & Weilert, E. "The Effects of Test-Wiseness Materials on Standardized Test Performance of Pre-School Disadvantaged Children," Paper presented at Feb. 1971 convention of The American Educational Research Association, New York.

Sassenrath, J.M. "Anxiety, Aptitude, Attitude and Achievement." Psychology in the School, 1967, 4, 341-346.

Slakter, M.J. "Generality of Risk Taking on Objective Examinations," Educational and Psychological Measurement, 1969, 29, 115-128.

Slakter, M.J. "Risk Taking on Objective Examinations." American Educational Research Journal, 1967, 4, 31-43.

Slakter, M.J., Koehler, R.A., Hampton, S.H. & Grennel, R.L. "Sex, Grade Level, and Risk Taking on Objective Examinations," Journal of Experimental Education, 1971, 39, 65-68.

Slakter, M.J., Koehler, R.A., & Hampton, S.H. "Grade Level, Sex, and Selected Aspects of Test-Wiseness," Journal of Educational Measurement, 1970, 7, 119-122.

Slakter, M.J., Koehler, R.A., & Hampton, S.H. "Learning Test-Wiseness By Programmed Texts," Journal of Educational Measurement, 1970, 7, 247-254.

Stanely, J.C. "Reliability," in Thorndike, R.L. (ed) Educational Measurement, Washington, D.C.: ACE, 1971, 356-442.

Stone, L.A. "Reliability of a Utility of Risk Measure," Psychological Reports, 1962, 10, 516.

Stricker, L.J. "Test-Wiseness on Personality Scales." Journal of Applied Psychology, 1969, 53, 1-18.

Thorndike, R.L. Personnel Selection: Test and Measurement Techniques. New York: John Wiley & Sons, 1949.

Vernon, P.E. "The Determinants of Reading Comprehension." Educational and Psychological Measurement, 1962, 22, 269-286.

Wahlstrom, M. & Boersma, F.J. "The Influence of Test-Wiseness Upon Achievement," Educational and Psychological Measurement, 1968, 28, 413-420.

Wevrick, L. "Response Set in a Multiple-Choice Test." Educational and Psychological Measurement, 1962, 22, 533-538.

ERIC Clearinghouse

JUN 6 1972

on Adult Education